13 December 2023

DCAT-AP for data spaces @EUROGEOGRAPHICS

DIGIT.B2 - Interoperability.

interoperable europe

*"Data spaces should foster an ecosystem (of companies, civil society and individuals) creating new products and services based on **more accessible data**."*
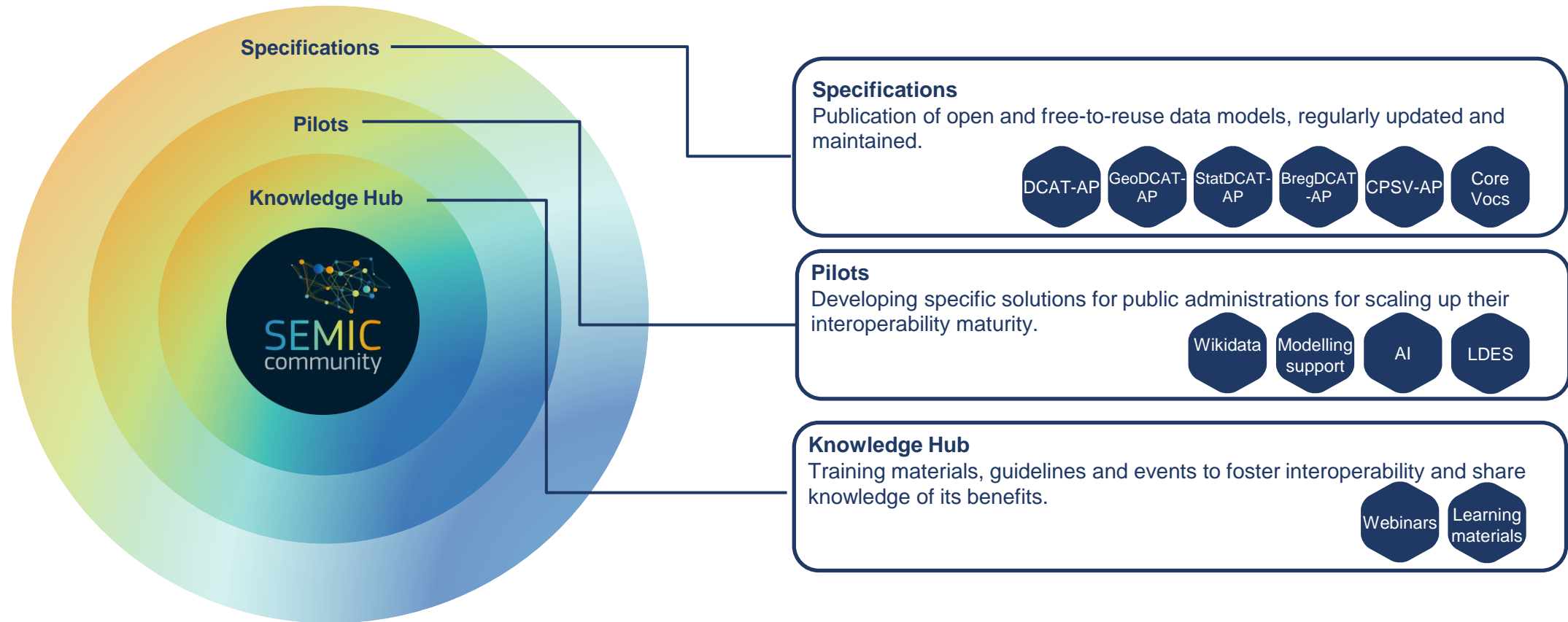A European strategy for data

Policy Context

> Europe's digital future will be enabled by a **data-driven economy** and the use of **Artificial Intelligence**, fully respecting EU values and regulations. The public sector also needs to become more data-driven; improve the capability of developing policies and services through the management, *__sharing and use of data__*.

> The [European Data Strategy](#) aims to create a single market for data through common **European data spaces** that benefit from *__common standards and interoperability protocols__*.

> Adopted by the European Commission in December 2022, the [Implementing Act on High-value Datasets](#) aims "*to make more publicly-funded information available for new information products and innovation, in particular in artificial intelligence*" (six categories: **geospatial**, earth observation and environment, meteorological, statistics, companies and mobility).

> The [Interoperable Europe Act](#) complements the EU data and digital policy landscape on *__data altruism__*. It implements interoperability by design and fosters the *__sharing and reuse of interoperable solutions__*.
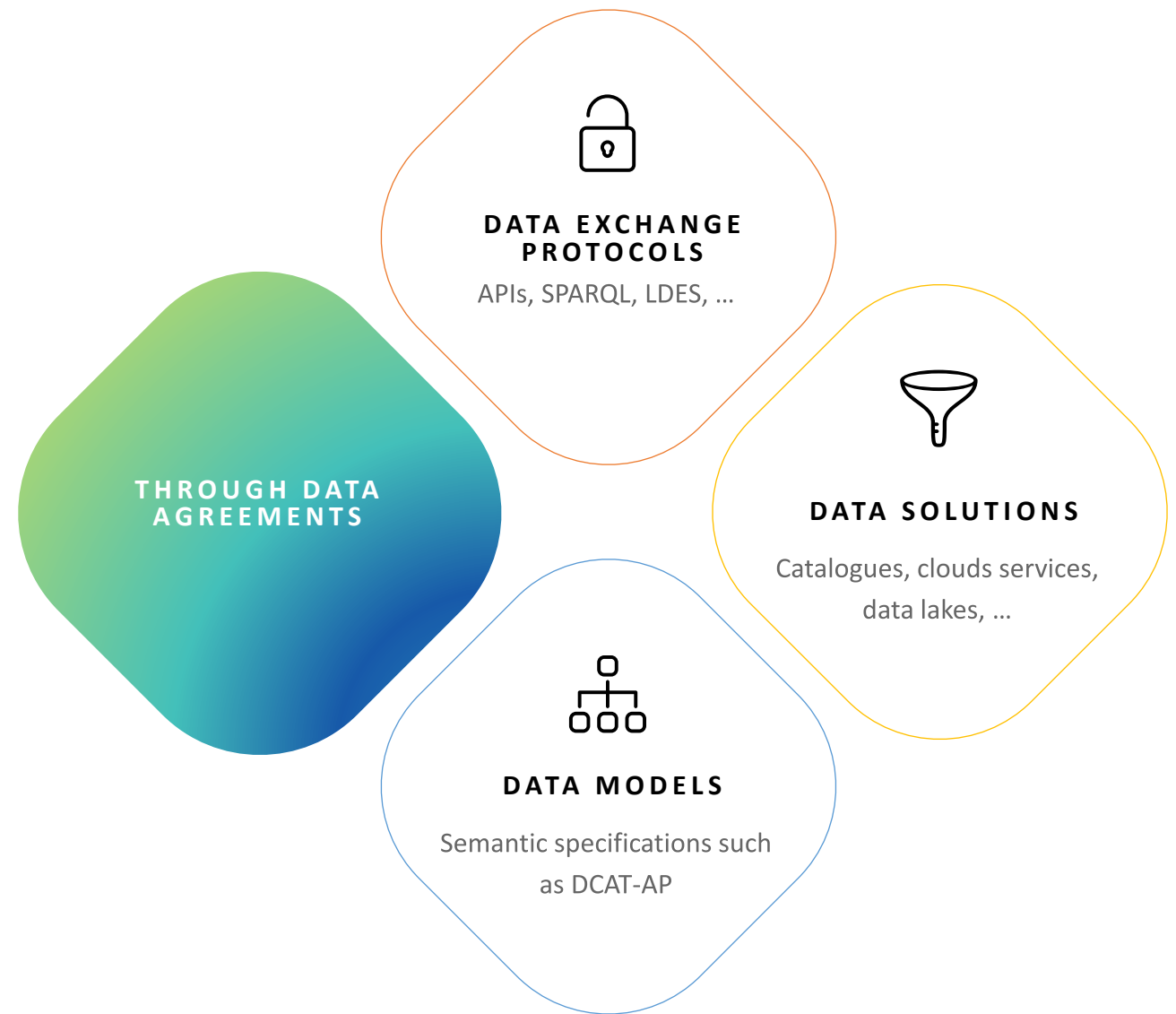
SEMIC

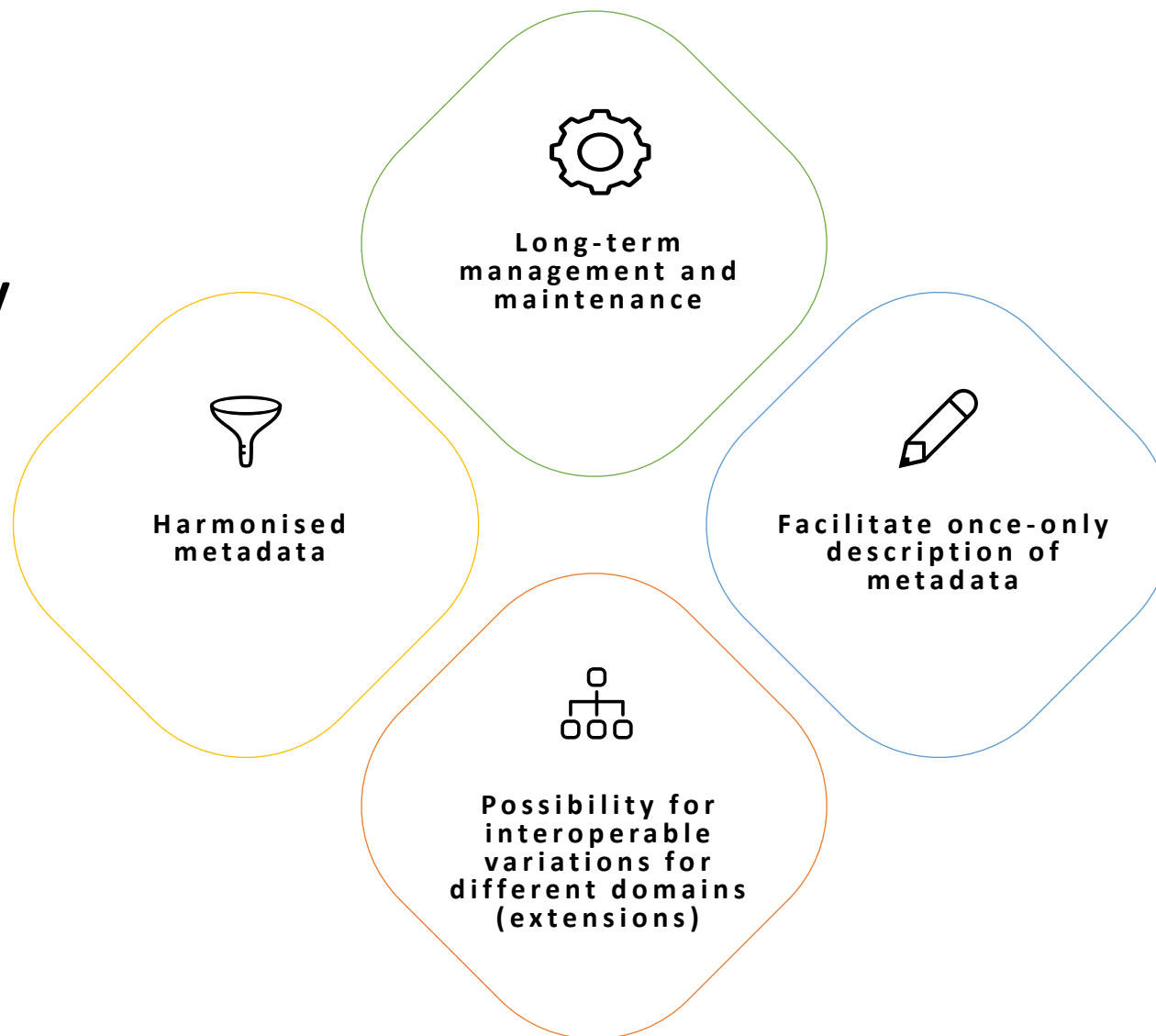SEMIC's goal is to deliver pragmatic support to help build an interoperable Europe.

Specifications

Pilots

Knowledge Hub

SEMIC community

Specifications
Publication of open and free-to-reuse data models, regularly updated and maintained.

DCAT-AP · GeoDCAT-AP · StatDCAT-AP · BregDCAT-AP · CPSV-AP · Core Vocs

Pilots
Developing specific solutions for public administrations for scaling up their interoperability maturity.

Wikidata · Modelling support · AI · LDES

Knowledge Hub
Training materials, guidelines and events to foster interoperability and share knowledge of its benefits.

Webinars · Learning materials

How to make Data more accessible?

European Data Strategy

THROUGH DATA AGREEMENTS

DATA EXCHANGE PROTOCOLS
APIs, SPARQL, LDES, …

DATA SOLUTIONS
Catalogues, clouds services, data lakes, …

DATA MODELS
Semantic specifications such as DCAT-AP

Interoperability for Dataset sharing

DCAT-AP

Long-term management and maintenance

Harmonised metadata

Facilitate once-only description of metadata

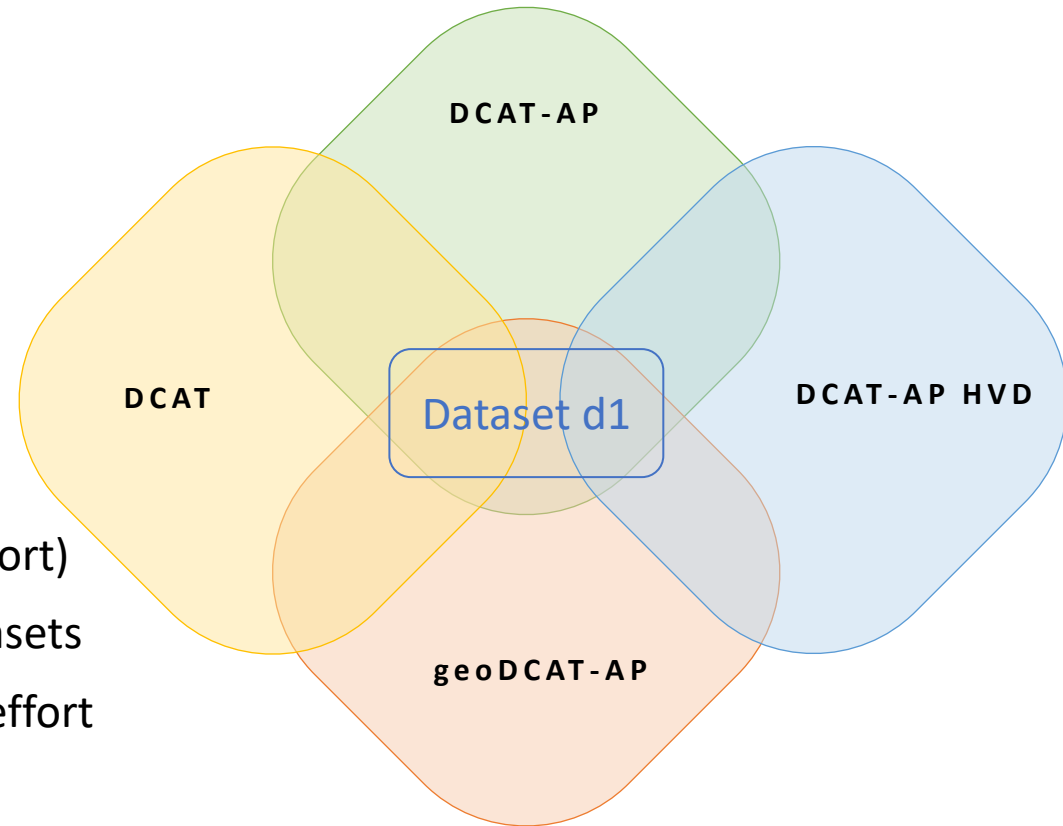Possibility for interoperable variations for different domains (extensions)

Multi-domain collaboration

Interoperable Profiles

- Concise (easiness to read, editorial effort)
- Once-only effort for publishers of datasets

resulting in acceptable implementation effort

DCAT-AP Profiles

A growing ecosystem

2012

DCAT 1

DCAT-AP
FOR
DATA PORTALS
IN EUROPE

DCAT-AP Profiles

A growing ecosystem

2015

DCAT 1

GEODCAT-AP FOR GEOSPATIAL DATASETS

DCAT-AP FOR DATA PORTALS IN EUROPE

STATDCAT-AP FOR STATISTICAL DATASETS

DCAT-AP Profiles

A growing ecosystem

2020

DCAT 2

GEODCAT-AP 2
FOR
GEOSPATIAL
DATASETS

DCAT-AP 2
FOR
DATA PORTALS
IN EUROPE

STATDCAT-AP
FOR
STATISTICAL
DATASETS

BRegDCAT-AP
FOR
BASE REGISTRIES

DCAT-AP Profiles

A growing ecosystem

2021

DCAT 2

GEODCAT-AP 2
FOR
GEOSPATIAL
DATASETS

DCAT-AP 2
FOR
DATA PORTALS
IN EUROPE

STATDCAT-AP
FOR
STATISTICAL
DATASETS

BRegDCAT-AP
2
FOR
BASE REGISTRIES

DCAT-AP in 2023

(only the activities by SEMIC)

Webinar Dataspaces 3/10/2023

Webinar 2 24 / 02 / 2023

Webinar 4 7/11/2023

DCAT-AP HVD FOR HIGH VALUE DATASETS

Public review

Webinar 1 02 / 02 / 2023

Webinar 3 30 / 03 / 2023

Start 18 / 07 / 2023

End 30 / 09 / 2023

Webinar 5 21 / 11 / 2023

DCAT-AP FOR DATA PORTALS IN EUROPE

Online interactions on https://github.com/SEMICeu/DCAT-AP

DCAT-AP

DCAT-AP

Major improvement trigged by alignment with DCAT 3.0

- Dataset Series (DCAT 3.0)

- From PDF to online html representation

- SEMIC Style Guide and Data Space profile alignment

4 webinars + 80 issues

Updated DCAT-AP specification based on provided feedback in GitHub and webinars.

To be released Jan 2024.

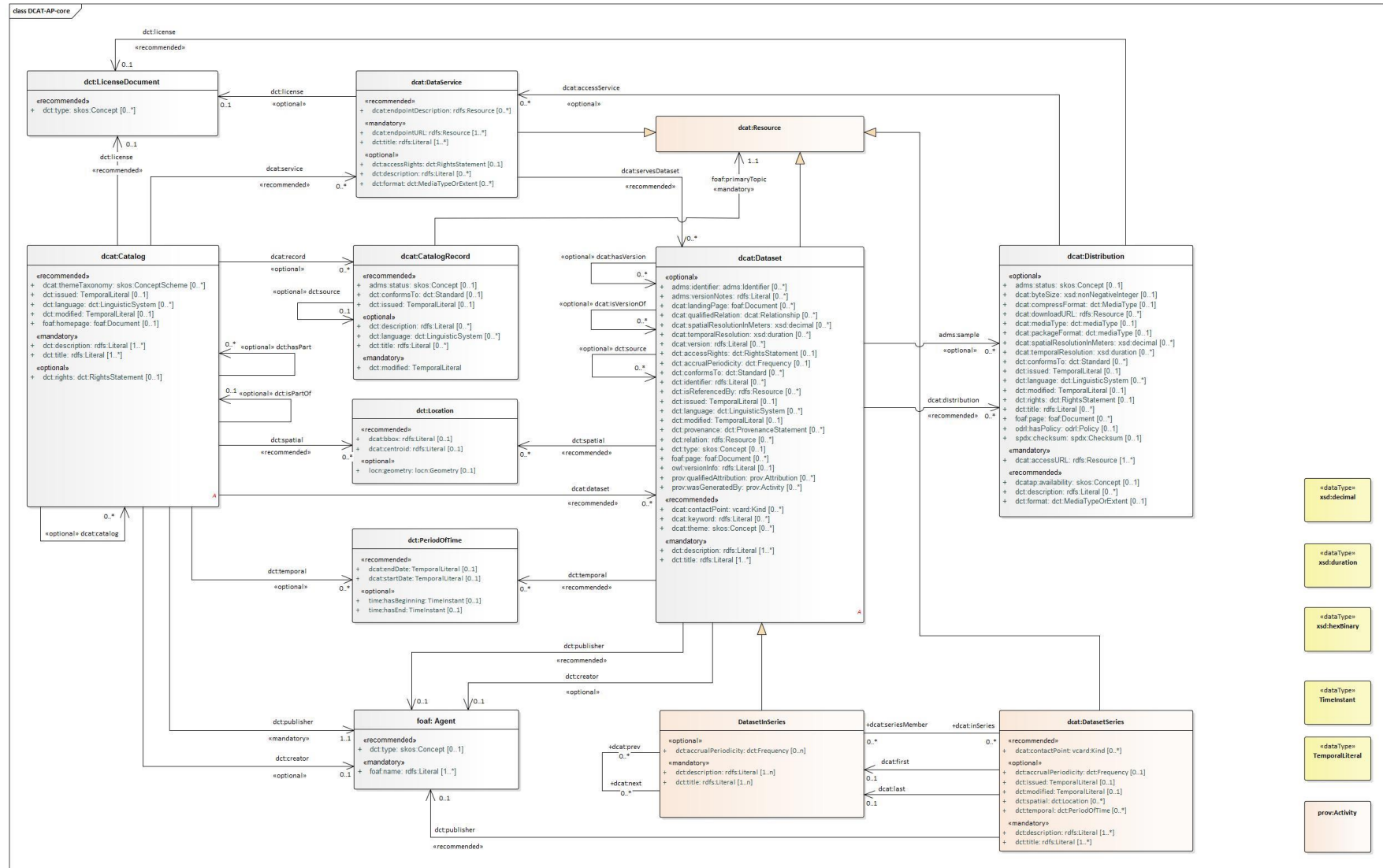DCAT-AP HVD as annex is compatible with previous release but also the next release DCAT-AP 3.0

https://semiceu.github.io/DCAT-AP/releases/3.0.0/

DCAT-AP

A great potential to document your datasets

SEMIC Style Guide for DCAT-AP Profile Alignment

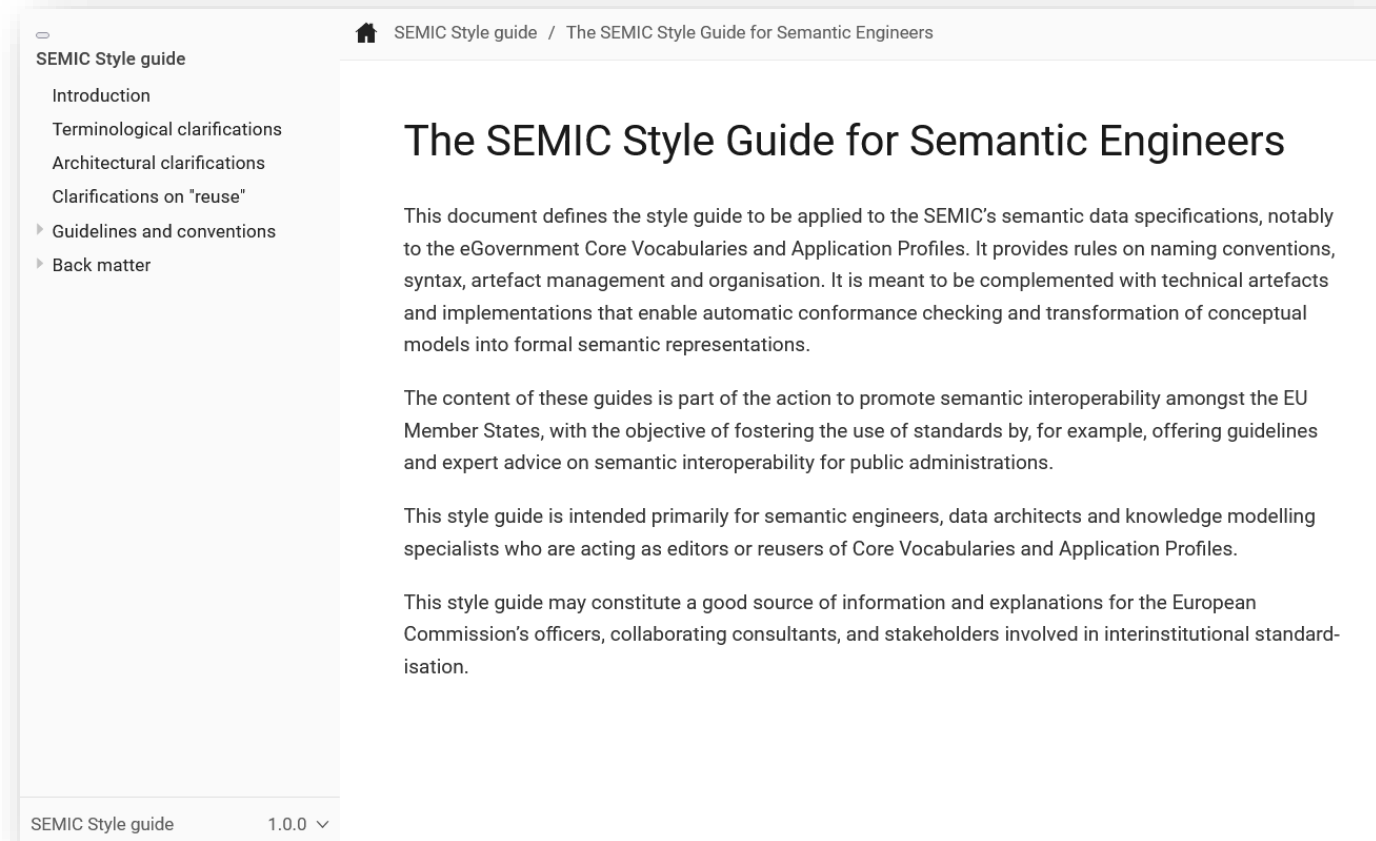Q? What are the rules to create good, interoperable data specifications.

Reuse

- as-is

- with terminological changes

- with semantical adaptions

Coherency & editorial effort

- internal coherent (all artifacts express the same knowledge)

- external reusable (artifacts ready for reuse)

SEMIC Style guide
Introduction
Terminological clarifications
Architectural clarifications
Clarifications on "reuse"
▸ Guidelines and conventions
▸ Back matter

SEMIC Style guide / The SEMIC Style Guide for Semantic Engineers

The SEMIC Style Guide for Semantic Engineers

This document defines the style guide to be applied to the SEMIC's semantic data specifications, notably to the eGovernment Core Vocabularies and Application Profiles. It provides rules on naming conventions, syntax, artefact management and organisation. It is meant to be complemented with technical artefacts and implementations that enable automatic conformance checking and transformation of conceptual models into formal semantic representations.

The content of these guides is part of the action to promote semantic interoperability amongst the EU Member States, with the objective of fostering the use of standards by, for example, offering guidelines and expert advice on semantic interoperability for public administrations.

This style guide is intended primarily for semantic engineers, data architects and knowledge modelling specialists who are acting as editors or reusers of Core Vocabularies and Application Profiles.

This style guide may constitute a good source of information and explanations for the European Commission's officers, collaborating consultants, and stakeholders involved in interinstitutional standardisation.

SEMIC Style guide   1.0.0 ⌄

https://semiceu.github.io/style-guide/1.0.0/index.html

SEMIC Style Guide for DCAT-AP Profile Alignment: Example

Q? As profile I want to express that each Dataset has a theme according to my codelist.

Reuse

- dcat:theme

- DCAT-AP already uses it, with another codelist

- For publishers, it should be an additional effort to extend the metadata of dataset published according to DCAT-AP with an additional value from another codelist.

  - Impact on software implementers should be acceptable (both editor as for data portal UI).

SEMIC Style Guide for DCAT-AP Profile Alignment: Example

Q? As profile I want to express that each Dataset has a theme according to my codelist.

Approach

- A new (mandatory) codelist is impactful

- Solution: a subproperty of dcat:theme with the new codelist as range requirement

- Impact:

  - It is DCAT compliant (through subproperty mapping a valid DCAT representation can be made)

  - It is DCAT-AP compliant (through distinct property the codelist requirements do not interfer)

  - It is easily implementable for software implementers

  - Validation rules are additional (no cross impact on artifacts from one profile on another)

DCAT-AP annex for
High Value Datasets

DCAT-AP for High Value Datasets

A mapping of the (generic) metadata requirements in HVD IR in DCAT-AP

(key elements: Dataset, Bulk Download, API, legal information)

Existing DCAT-AP datasets with a minimum of effort extended to conform to DCAT-AP HVD

As cross domain metadata it can be used to create coherent HVD reporting

To be used by domains to assess their current metadata practices

3 webinars + 28 issues on GitHub, joined collaboration with PSI and DCAT-AP experts (CNECT and SEMIC)

The first release is launched **on 14 dec 2023**.

https://semiceu.github.io/DCAT-AP/releases/2.2.0-hvd/

DCAT-AP for High Value Datasets

Additional constraints on core DCAT-AP classes

DCAT-AP for High Value Datasets: Quick Overview (1/3)

1. Mapping terminology on core classes of DCAT-AP:

   1. High Value Dataset = dcat:Dataset

   2. API = dcat:DataService

   3. Bulk Download = dcat:Distribution

2. Identification HVD

   Use (new) property *applicable legislation* with ELI HVD IR (http://data.europa.eu/eli/reg_impl/2023/138/oj)

3. Categorisation HVD

   Use (new) property *HVD category* is mandatory for Dataset and Distribution with controlled vocabulary (http://data.europa.eu/bna/asd487ae75)
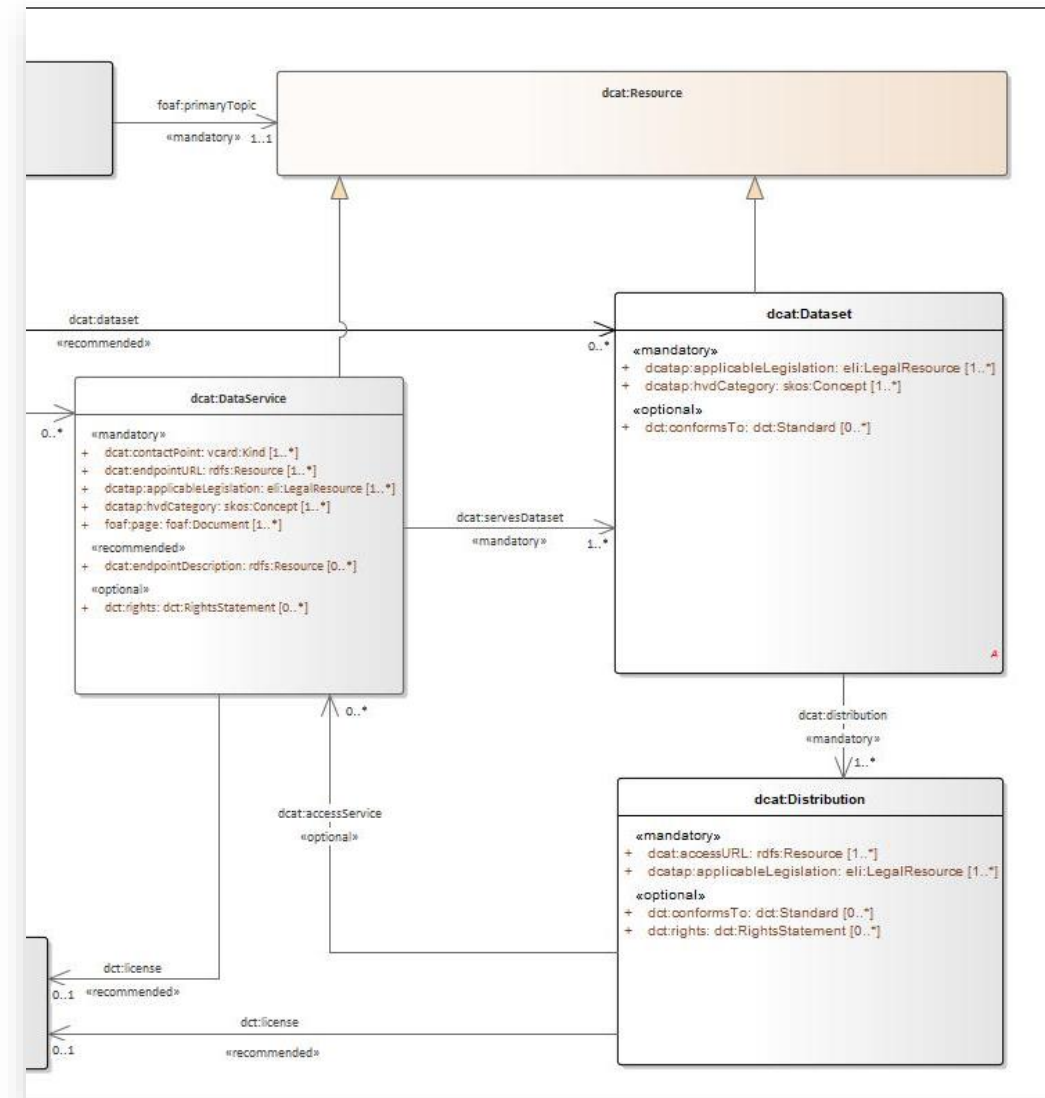
DCAT-AP for High Value Datasets: Quick Overview (2/3)

1. Use persistent identifiers
2. Express legal information
   1. in a human and machine readable approach
   2. As least as permissive as CC BY 4.0
   3. Approach expressed in a flow diagram
3. Provide contact details for APIs
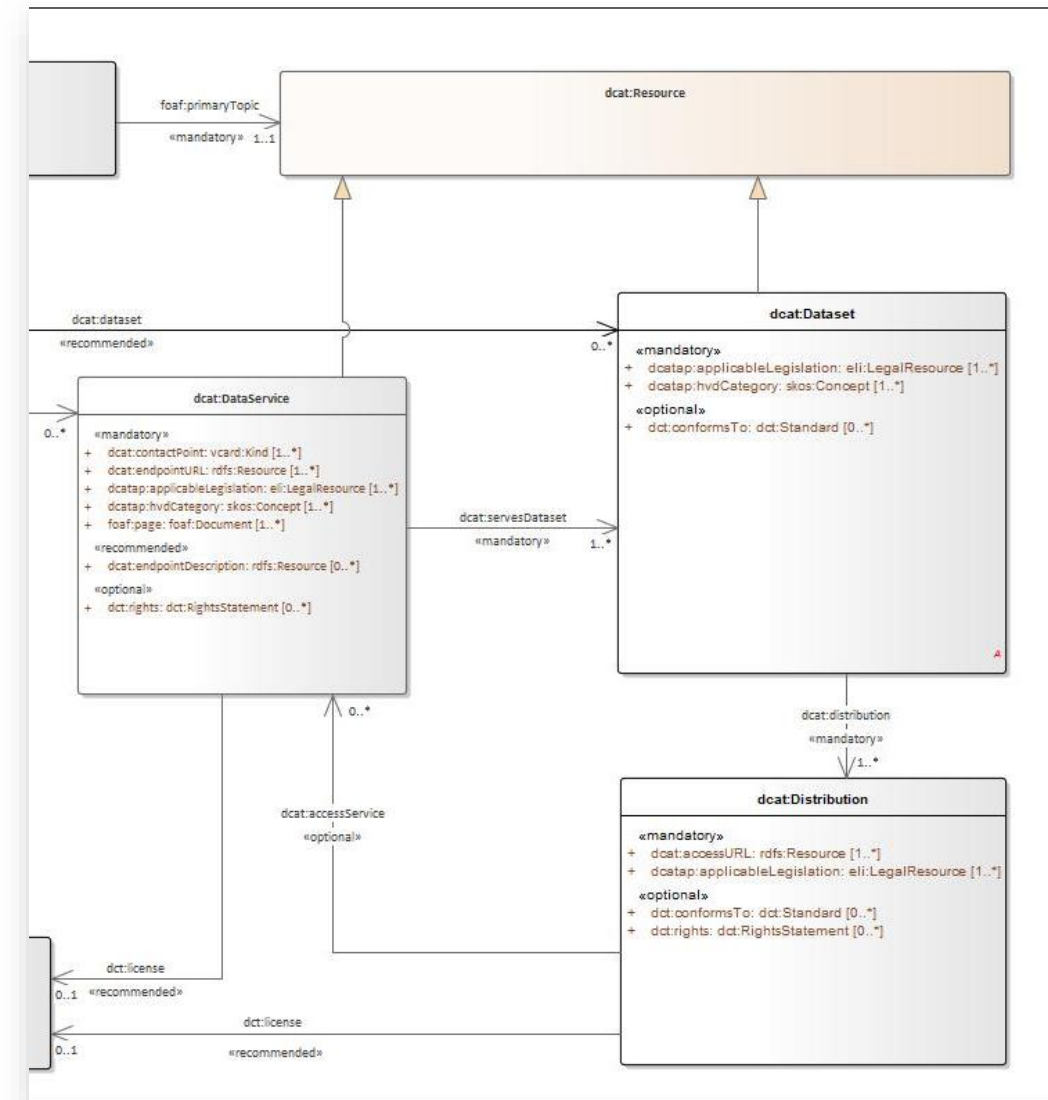
DCAT-AP for High Value Datasets: Quick Overview (3/3)

1. DCAT-AP HVD is an annex of DCAT-AP

   = additional constraints in case the Dataset under consideration is subject to the HVD IR

   a. References to Webinars

   b. Cross references to DCAT

   c. Cross references to DCAT-AP for each property

      A (as is), **E** (with additional constraints), **P** (profile specific)

   Additional information to highlight the relationship with DCAT-AP

   d. Additional information such as examples and validation support

2. DCAT-AP HVD is compatible with DCAT-AP 2.x but also with DCAT-AP 3.0

---

§ 7.5 Dataset

Definition
A conceptual entity that represents the information published.

Reference in DCAT
Link

Subclass of
Catalogued Resource

Properties
For this entity the following properties are defined: applicable legislation , conforms to , contact point , dataset distribution , HVD Category .

Property	Range	Card	Definition	Usage	DCAT	Reuse
applicable legislation	Legal Resource	1..*	The legislation that mandates the creation or management of the Dataset.	For HVD the value must include the ELI http://data.europa.eu /eli/reg_impl/2023/138/oj. As multiple legislations may apply to the resource the maximum cardinality is not limited.		P
conforms to	Standard	0..*	An implementing rule or other specification.	The provided information should enable to the verification whether the detailed information requirements by the HVD is satisfied. For more usage suggestions see section on specific data requirements.	Link	A
contact point	Kind	0..*	Contact information that can be used for sending comments about the Dataset.		Link	A
				The HVD IR is a quality		

DCAT-AP for geographical data

GeoDCAT-AP

Status: aligned with DCAT 2 since end 2020

Meant to provide a DCAT-AP compliant representation for the set of metadata elements
included in INSPIRE metadata
(ISO 19115:2003 and ISO 19119)

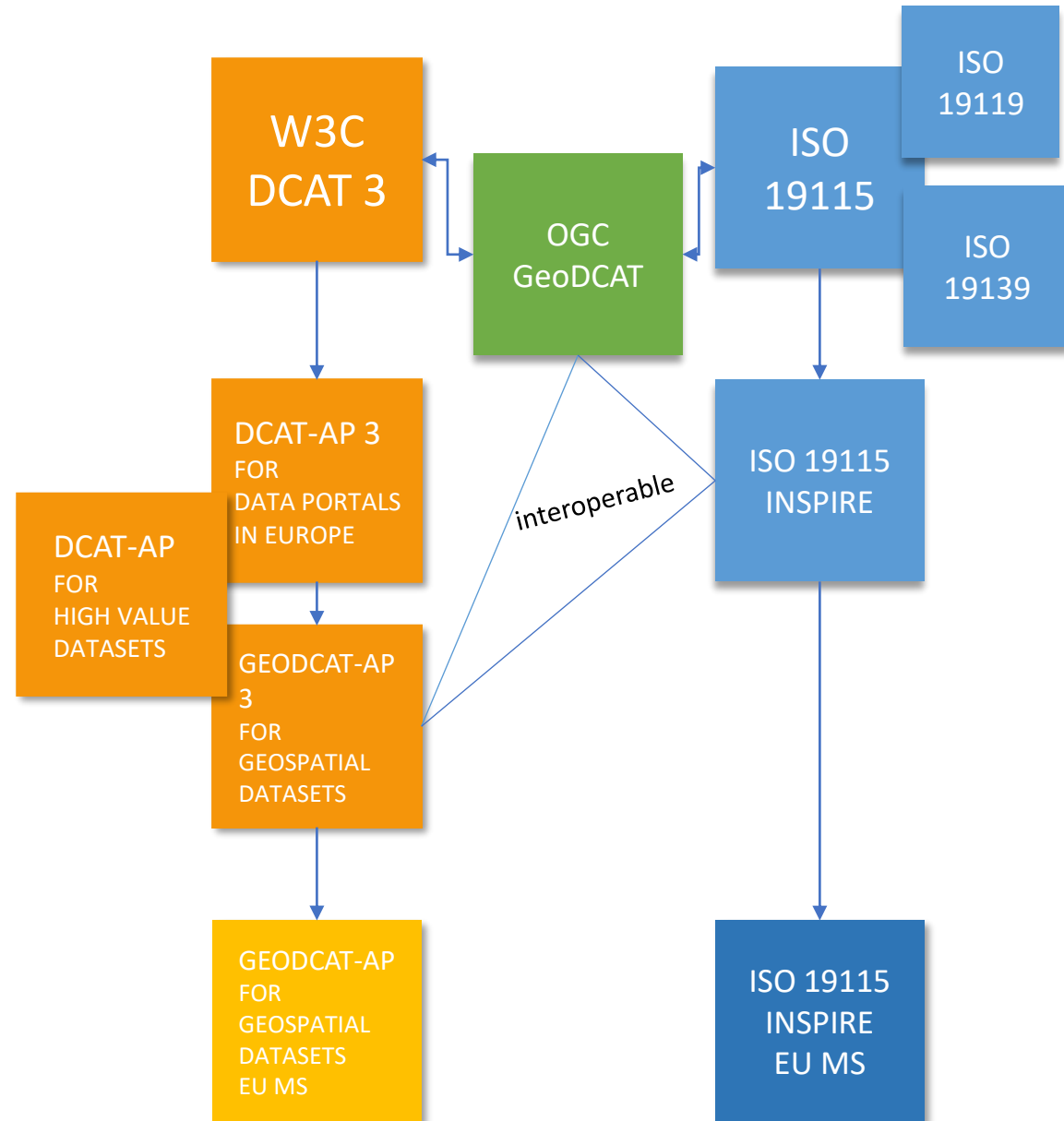A INSPIRE-to-RDF mapping ensures metadata interoperability and sharing

Upcoming: alignment with DCAT-AP 3 in 2024

https://semiceu.github.io/GeoDCAT-AP/releases/2.0.0/

GeoDCAT-AP: ecosystem

Interoperability between specifications

(even cross Standardisation Bodies)

GeoDCAT-AP: ecosystem

Interoperability between specifications
(even cross Standardisation Bodies)

Supporting tools:

- XSLT (ISO -> GeoDCAT-AP RDF)
  - Use in your local system
  - Deployed as API
  - For a whole catalogue CSW-4-web
- Supportive URIs:
  - EPSG codelist
  - Geometry URIs
  - …

Supporting tools: GeoDCAT-AP XSLT

XML

INSPIRE / ISO 19115

GeoDCAT-AP XSLT

RDF/XML
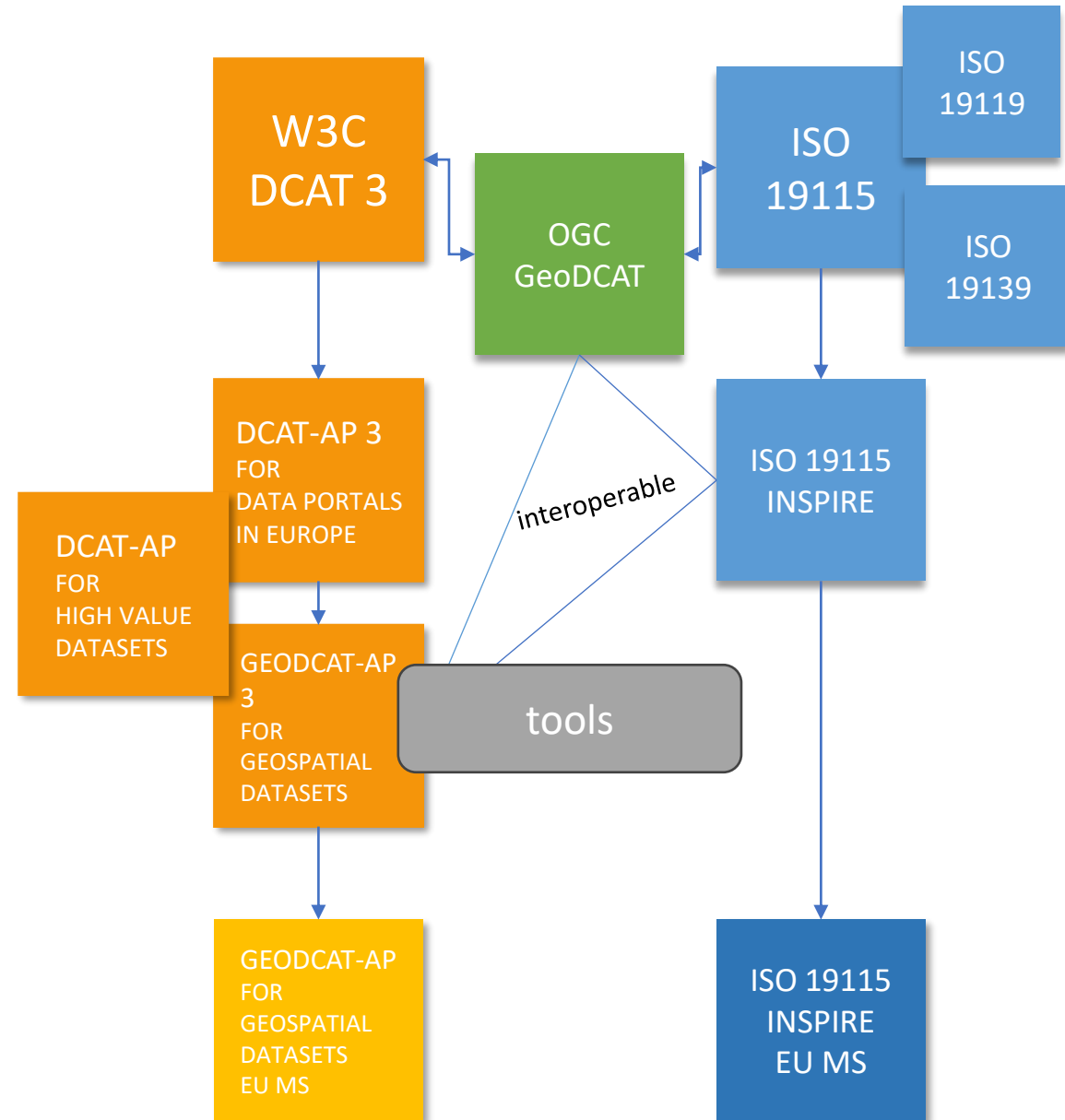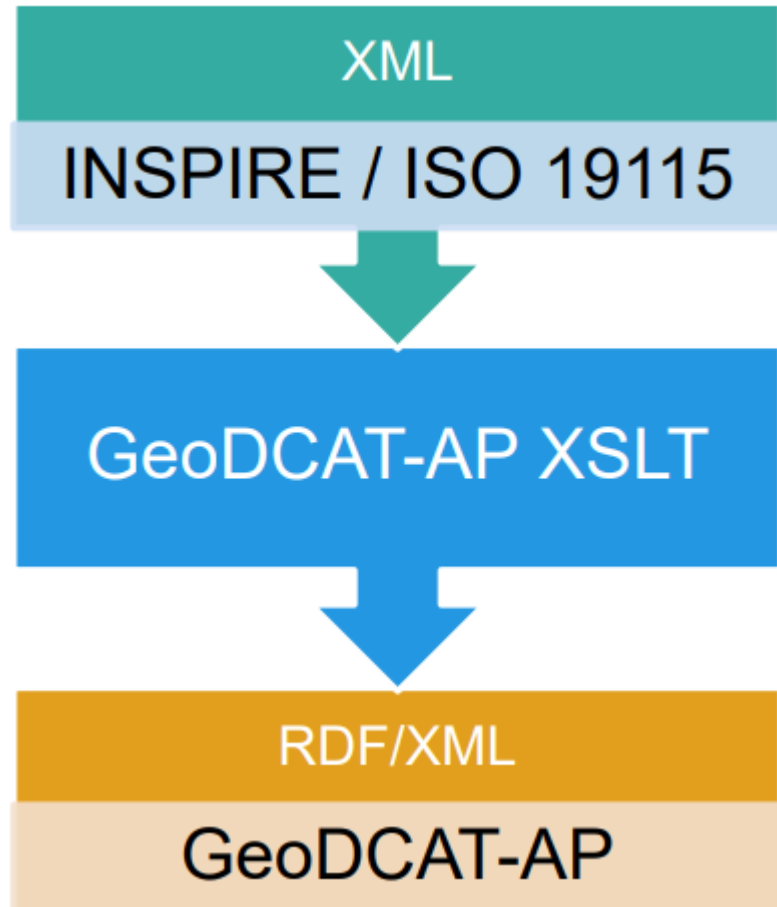
GeoDCAT-AP

- Reference implementation of the mappings defined in Geo-DCAT-AP

- Converts INSPIRE / ISO 19115 into Geo-DCAT-AP

- Can run in any language and tool supporting XSLT

https://github.com/SEMICeu/iso-19139-to-dcat-ap

Supporting tools: GeoDCAT-AP API

XML

INSPIRE / ISO 19115

GeoDCAT-AP XSLT

GeoDCAT-AP API

RDF/XML, Turtle, JSON-LD
…
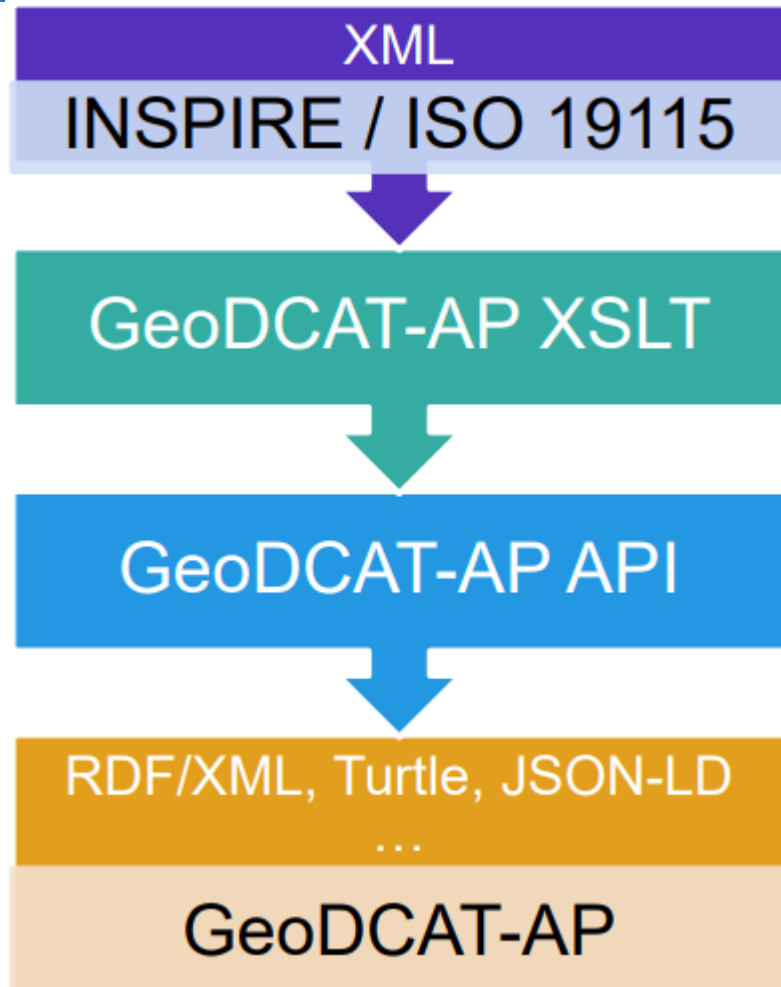
GeoDCAT-AP

- Given a CSW endpoint, extract all dataset descriptions according to the XSLT transformation in GeoDCAT-AP.

- Uses the Geo-DCAT-AP XSLT

- multiple RDF serializations output

http://geodcat-ap.semic.eu/api/

Supporting tools: GeoDCAT-AP API

GeoDCAT-AP API
ISO 19139 records in RDF

Output Schema : GeoDCAT-AP ⌄   **Transform**

https://sdi.eea.europa.eu/catalogue/srv/eng/csw?request=GetRecords&service=CSW&ver

Output format : HTML+RDFa ⌄

Usage notes

Copy & paste the URL of a file or of a CSW request returning ISO 19139 records.

Supported CSW request types: GetRecords, GetRecordById.

Supported CSW output schema: http://www.isotc211.org/2005/gmd

NB: The current version of the API supports only CSW calls using the GET HTTP method.

A description of the GeoDCAT-AP API is available on the dedicated GitHub repository.

http://geodcat-ap.semic.eu/api/

Supporting tools: CSW-4 Web

- Publish the content of a CSW endpoints in a web-friendly way, and enabling the exploration of its content without the need of specific client applications

- Uses an extended and ad-hoc version of Geo-DCAT-AP XLST & API
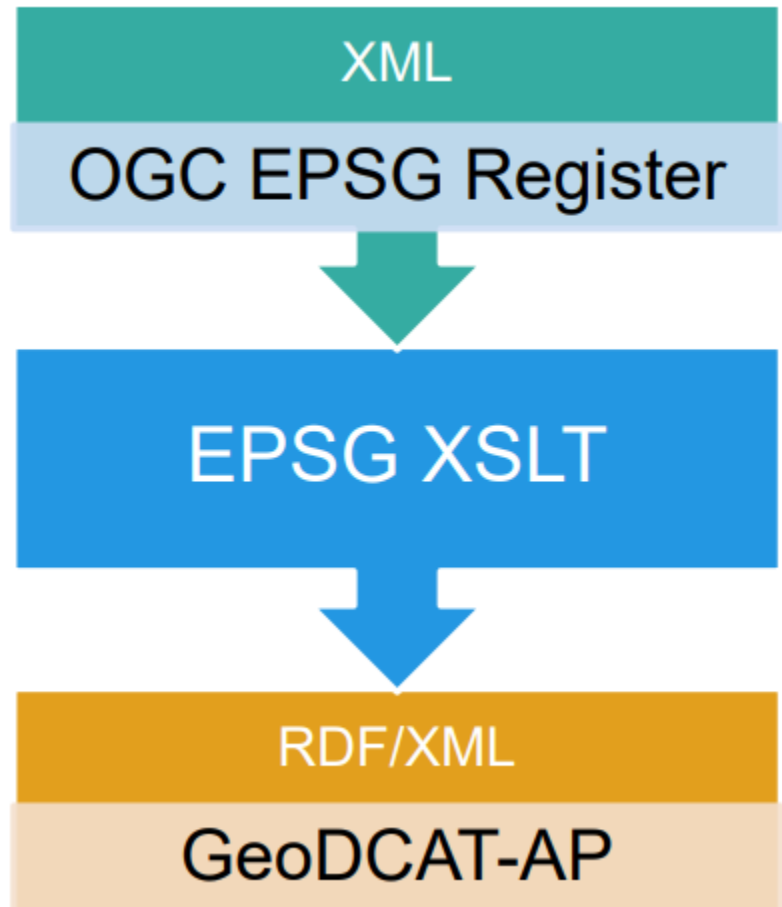
http://geodcat-ap.semic.eu/csw-4-web/
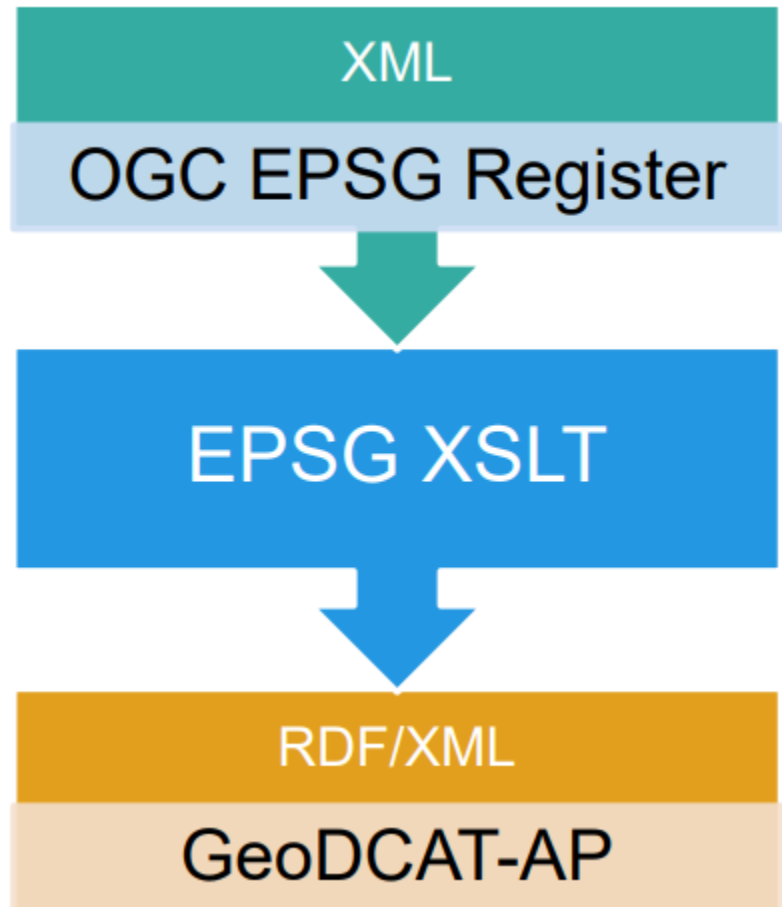
Other tools: EPSG-XSLT

XML

OGC EPSG Register

EPSG XSLT

RDF/XML

GeoDCAT-AP

- Converts entries from the OGC EPSG register of CRS into a GEO-DCAT-AP conformant representation

- Can be run in any language and tool supporting XLST

http://github.com/SEMICeu/epsg-to-rdf/

Other tools: EPSG-XSLT

- Converts entries from the OGC EPSG register of CRS into a GEO-DCAT-AP conformant representation

- Can be run in any language and tool supporting XLST

http://github.com/SEMICeu/epsg-to-rdf/

Other tools

GeoIRI
HTTP URIs for geometries

Geometry (WKT)   ⓘ EPSG : 4326   **Get GeoIRI**

MULTIPOLYGON(((40 40, 20 45, 45 30, 40 40)),((20 35, 45 20, 30 5, 10 10, 10 30, 20 35),(30 20, 20 25, 20 15, 30 20)))

NB: The axis order in the WKT-encoded geometry must be longitude / latitude, irrespective of the selected coordinate reference system.

Powered by GeoIRI

This work is jointly supported by the ARe3NA and SEMIC Actions of the EU ISA Programme.

ARe3NA SEMIC ISA

Measurement resolution
HTTP URIs for measurement resolution

Scale: 1:100

URI: http://dcat-ap.semic.eu/id/resolution/scale/100
Type: Quality Measurement
Raw data: RDF/XML | N-Triples | N3 | Turtle | JSON-LD |
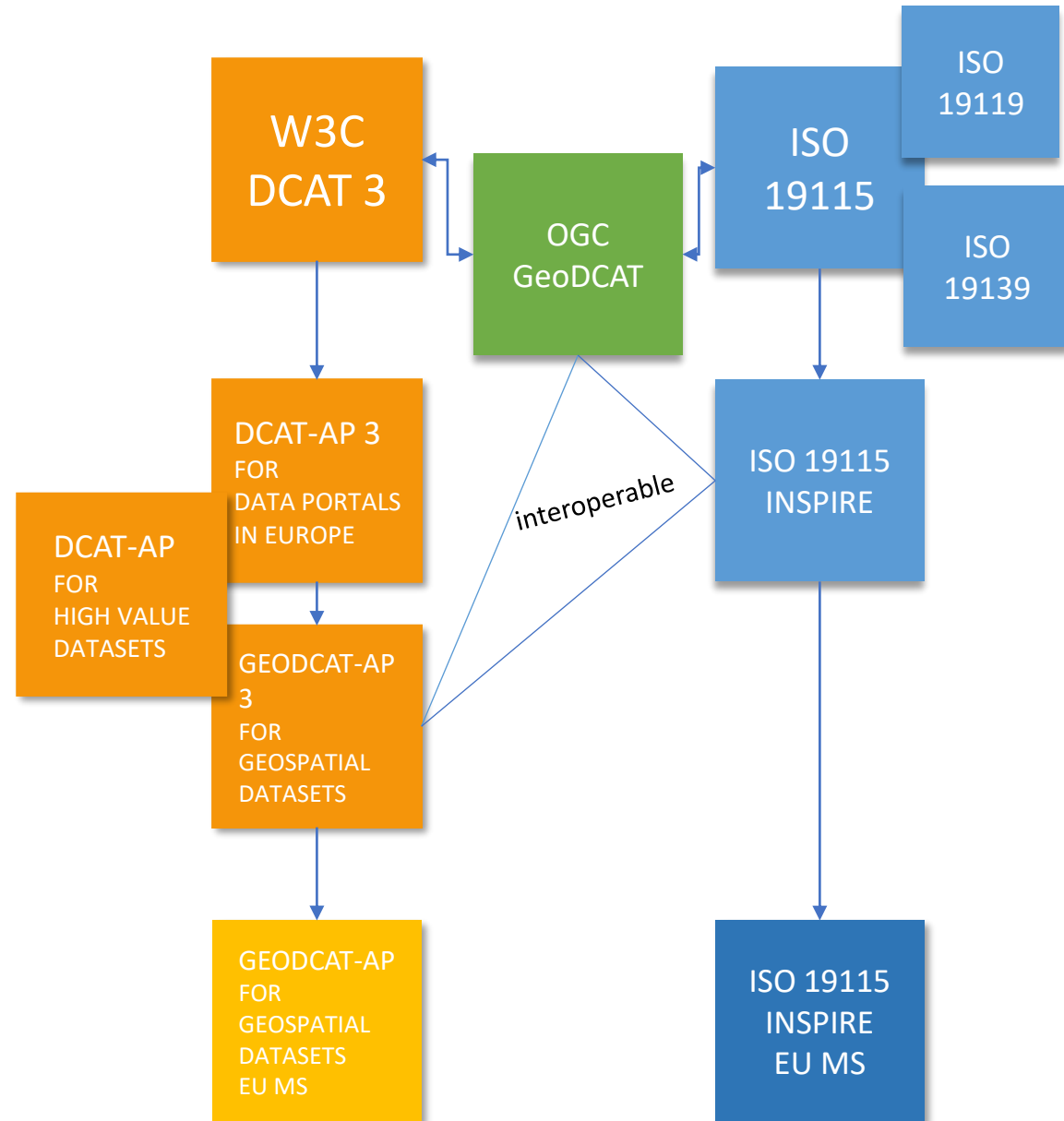
Properties

rdfs:label	Spatial resolution (scale): 1:100
rdf:type	dqv:QualityMeasurement
dqv:isMeasurementOf	geodcatap:SpatialResolutionAsScale
dqv:value	0.0100000000000000000208167

Create URIs for geometries

https://geodcat-ap.semic.eu/geoiri/

Other supportive URIs

https://geodcat-ap.semic.eu/id/resolution/scale/100

GeoDCAT-AP 3 in 2024

Your contributions are welcome.

Thank you

interoperable
europe

innovation ∞ govtech ∞ community

Stay in
touch

(@InteroperableEU) / Twitter

Interoperable Europe - YouTube

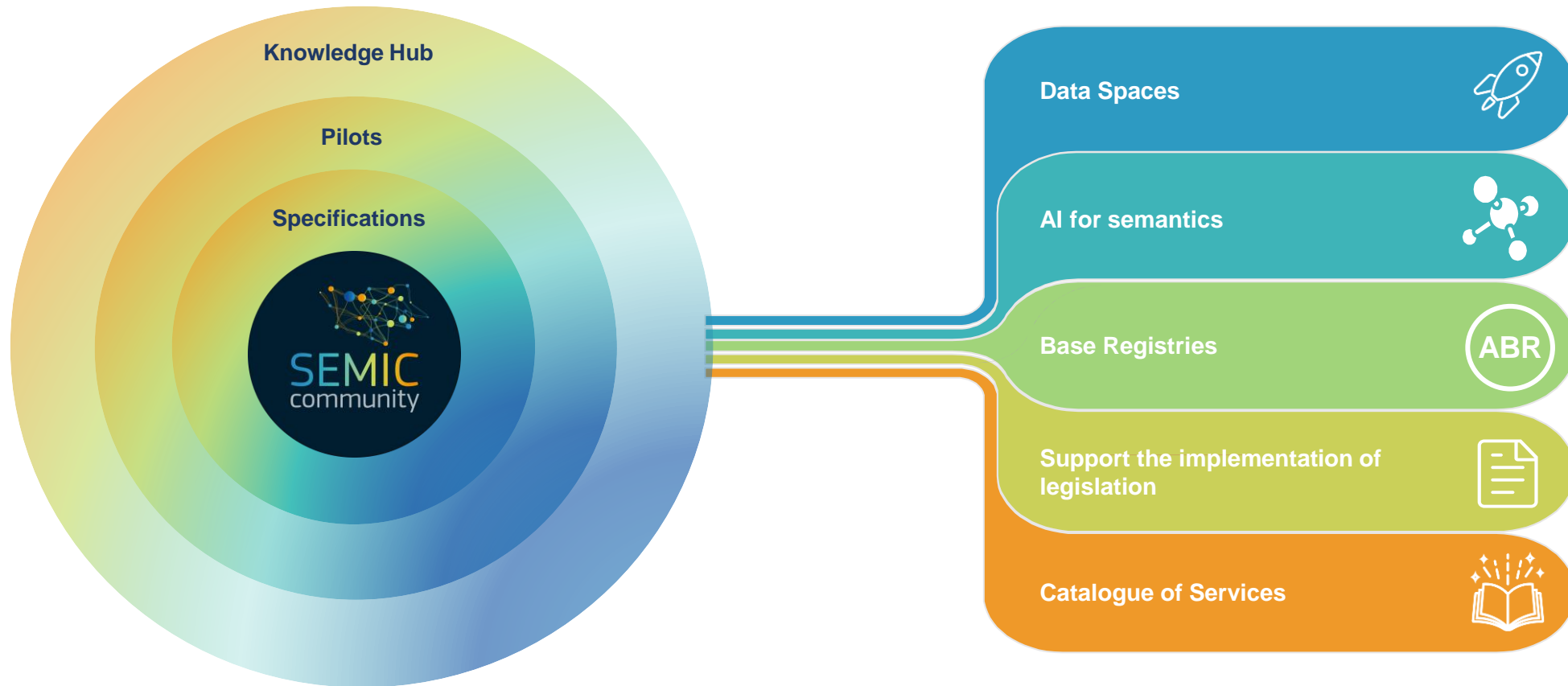Interoperable Europe | LinkedIn

DIGIT-INTEROPERABILITY@ec.europa.eu

https://joinup.ec.europa.eu/collection/interoperable-europe/interoperable-europe

Dataset sharing: interoperability across data spaces

SEMIC Focus Areas

Knowledge Hub

Pilots

Specifications

SEMIC community

Data Spaces

AI for semantics

Base Registries

ABR

Support the implementation of legislation

Catalogue of Services

ACCESSIBLE

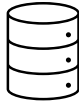FINDABLE

Data must be **FAIR**

INTEROPERABLE

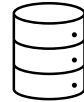REUSABLE

Example of Data Space Solutions

Agency for maintenance of roads

Publishes the road network in the area A.

Agency of environment

Publishes the emission of small dust particles in the area A.

Agency of health

Publishes the absence due to illness in the area A.

Combining data across data spaces

As a researcher I want to find the impact of road related emissions on work absence…

Findability of Catalogues

If there are three different portals…
How does the researcher find them?

Understanding dataset descriptions

If portals could share the metadata description…
Is the metadata compatible?

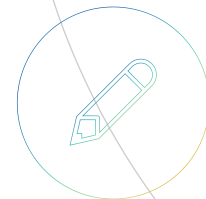# How do we enable interoperability…

across data spaces and
the wider semantic web, while

saving costs, time, and

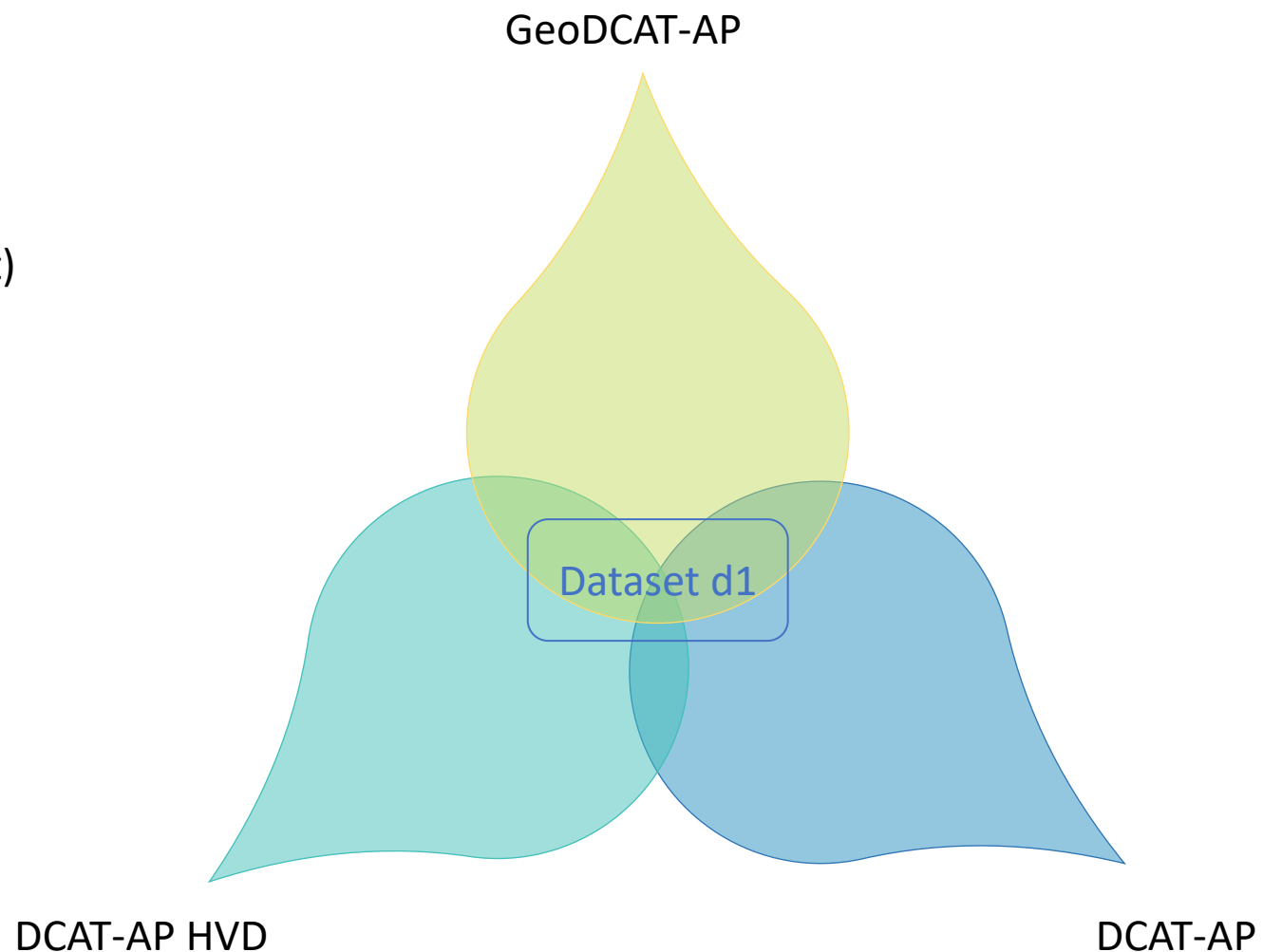reusing what is already there as
much as possible?

Interoperable Europe Act

Linking public services, supporting public policies and building public benefits

Multi-domain collaboration

- Conciseness (easiness to read, editorial effort)

- Implementation effort

- Once-only effort for publishers of datasets

GeoDCAT-AP

Dataset d1

DCAT-AP HVD

DCAT-AP

GeoDCAT-AP and INSPIRE

- Agree upon a common RDF representation
- RDF is increasingly being used as an alternative representation of INSPIRE metadata
  - Without a harmonised INSPIRE-to-RDF mapping, metadata interoperability is lost
- Facilitate cross-sector sharing of INSPIRE metadata
  - INSPIRE metadata are already being harvested by and published in cross-domain data catalogues at the national and/or regional level
  - INSPIRE metadata are harvested and published also on the European Data Portal, which uses DCAT-AP as a metadata interchange format

GeoDCAT-AP, INSPIRE and ISO Standards

- GeoDCAT-AP is meant to provide a DCAT-AP compliant representation for the set of metadata elements included in INSPIRE metadata

- The core profile of ISO 19115:2003

- The GeoDCAT-AP specification does not replace the INSPIRE Metadata Regulation nor the INSPIRE Metadata Technical Guidelines based on ISO 19115:2003 and ISO 19119

- Its purpose is to give owners of geospatial metadata the possibility to achieve more by providing an additional RDF syntax binding

- Its basic use case is to make spatial datasets, data series, and services searchable on general data portals, thereby making geospatial information better searchable across borders and sectors